AFRL-SN-RS-TR-2004-358
In-House Report
December 2004

KNOWLEDGE-BASED GENERALIZED LIKELIHOOD RATIO TEST (KB-GLRT): EXPLOITING KNOWLEDGE OF THE CLUTTER RIDGE IN AIRBORNE RADAR

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AIR FORCE RESEARCH LABORATORY
SENSORS DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

STINFO FINAL REPORT

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-SN-RS-TR-2004-358 has been reviewed and is approved for publication.

APPROVED: /s/

GERARD J. GENELLO, Jr., Chief
RadarSignalProcessingBranch

FOR THE DIRECTOR: /s/

RICHARD G. SHAUGHNESSY, Chief
Rome Operations Office
Sensors Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	DECEMBER 2004	In-House Oct 04 – Nov 04

4. TITLE AND SUBTITLE
KNOWLEDGE-BASED GENERALIZED LIKELIHOOD RATIO TEST (KB-GLRT): EXPLOITING KNOWLEDGE OF THE CLUTTER RIDGE IN AIRBORNE RADAR

5. FUNDING NUMBERS
C - N/A
PE - 62602F
PR - 517R
TA - 05
WU - 00

6. AUTHOR(S)
Antonio De Maio,
Alfonso Farina,
and Michael Wicks

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
DIET
Via Claudio 21, 80125
Napoli Italy

Alenia Marconi Systems
Via Tiburtina Km.12.4
00131 Roma Italy

AFRL/SN
26 Electronic Parkway
Rome New York 13441-4514

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/SN
26 Electronic Parkway
Rome New York 13441-4514

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
AFRL-SN-RS-TR-2004-358

11. SUPPLEMENTARY NOTES

AFRL Project Engineer: Michael C. Wicks/SN/(315) 330-2556/ Michael.Wicks@rl.af.mil

12a. DISTRIBUTION / AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

In this paper we address knowledge-based radar detection for STAP applications. To this end we exploit, at the design stage, the characteristic structure of the clutter ridge and devise two decision rules according to the Generalized Likelihood Ratio Test (GLRT) and the two-step GLRT criteria. We first focus on the case of a clutter ridge with integer slope and then discuss the more general framework of a non-integer slope parameter. With reference to this last case we only provide approximate GLRT detectors due to the analytical difficulties connected with the exact solution of the problem.

The performance analysis shows that the new knowledge-based systems achieve a performance level very close to the optimum detector which assumes the perfect knowledge of the clutter covariance matrix and can outperform some previously proposed adaptive detection schemes.

14. SUBJECT TERMS
Radar, Signal Processing

15. NUMBER OF PAGES
26

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	

Table of Contents

List of Figures

1 Introduction

Space-Time Adaptive Processing (STAP) is a technique used to reject interference [1, 2], offering the potential to improve airborne radar performance in several adverse cases such as low-velocity target detection as well as that of weak targets obscured by sidelobe clutter. It is based on a filter that adaptively processes the radar data received from L different spatial channels (sub-arrays or elements of a phased array) and N pulses in the coherent processing interval.

In [3] the optimum linear processor which maximizes the output signal to clutter/interference plus noise power ratio is derived under the assumption of disturbance with known covariance matrix. However, since the clutter/interference spectral properties are not known a priori and may also vary both in time and space, a structure which adaptively estimates the weights of the aforementioned filter has been proposed and assessed in [4]. It is based on a set of K secondary data collected from range gates spatially close to the one under test and sharing the same covariance. Besides, in order to achieve an effective clutter/interference suppression, a large number of range cells ($2NL$ for achieving an average performance within 3 dB of the optimal value) sharing the same covariance matrix is required. This is an important limitation since in real environments the number of data in which the clutter is homogeneous (often referred to as sample support) might be very limited.

In order to circumvent this drawback several techniques have been proposed in the recent past. They largely fall into four distinct classes:

1. Minimal sample support methods which include reduced-rank STAP [2], chapters 5,6, and [9], multistage Wiener filtering techniques [5], joint-domain localized algorithms [6], least squares space-time filters exploiting the property of finite correlation length [2], chapter [7], principal components techniques [7], pre/post-Doppler STAP [1] based also on a partial *a-priori* knowledge of the disturbance scenarios, methods exploiting structural information[1] about the disturbance covariance matrix [8, 9, 10,11,12], and, finally, covariance tapers [7].
2. Knowledge-aided STAP employing intelligent training, namely suitable techniques for a careful selection of the secondary data [13], or knowledge-based criteria [14, 15], which resort to *a-priori* information concerning the surrounding environment.
3. Higher-order methods which include non-linear filters [16, 17, 18] to suppress the clutter ridge and approximation by first-order Taylor series of the non-stationarity.
4. Deterministic (single shot) techniques [19].

In this paper we propose a knowledge-based approach for STAP which exploits the special structure of the clutter Power Spectral Density (PSD) [20]. Indeed it is well known that in the absence of Intrinsic Clutter Motion (ICM) and neglecting velocity misalignment the clutter ridge (namely the locus in the Doppler-angle domain where the clutter PSD is sensibly different from zero) is composed of a finite width line (Doppler unambiguous clutter) or of multiple finite width lines (Doppler ambiguous clutter). We use at the design stage this *a-priori* information which

[1] These methods are also known as parametric or model-based techniques.

for the case of integer clutter ridge slope β, as shown in Section 3, is equivalent to assuming the knowledge of the clutter covariance matrix subspace.

In order to synthesize the receivers we resort to the Generalized Likelihood Ratio Test (GLRT) and the two-step GLRT (2S-GLRT) criteria, and derive two new decision rules referred to as the Knowledge-Based GLRT (KB-GLRT) and the Knowledge-Based 2S- GLRT (KB-2S-GLRT) respectively. Actually, the quoted derivations refer to the case of integer β. Hence, in order to deal with the most general framework of non-integer β, we develop two approximate versions of the aforementioned receivers and discuss the possible exploitation of further *a-priori* information concerning the clutter covariance subspace.

The paper is organized as follows. In Section 2 we formulate the problem and in Section 3 we provide some insights about the clutter covariance matrix. In Sections 4 and 5 we devise the KB-GLRT and the KB-2S-GLRT respectively. The performance of the new receivers is presented in Section 6 also in comparison with several previously devised STAP detectors. Section 7 is devoted to the generalization of the novel decision rules to the case of a non-integer clutter ridge slope. Finally, conclusions are drawn in Section 8.

2 Problem Foundation

Data are collected by a narrowband antenna array with L spatial channels which, for simplicity, we assume colinear, omnidirectional, and equally spaced with spacing d. Each channel receives N echoes corresponding to the returns of a coherent pulse train composed of N pulses with a pulse repetition time T_p. Denote by x_i, $i = 1; : : : ;N$, the L-dimensional column vector of the snapshots along the spatial channels and let

$$\boldsymbol{r} = \left[\boldsymbol{x}_1^T,\ \boldsymbol{x}_2^T, \ldots,\ \boldsymbol{x}_N^T \right]^T \qquad (NL - \text{dimensional}) \tag{1}$$

with $(.)^T$ denoting the transpose operator.

Under the hypothesis H_0, namely target absence, the vector r contains disturbance only, i.e.

$$\boldsymbol{r} = \boldsymbol{d}, \tag{2}$$

where d denotes the overall disturbance accounting for both clutter and thermal noise.

Under H_1, r contains also a target component, i.e.

$$\boldsymbol{r} = \alpha \boldsymbol{p} + \boldsymbol{d}, \tag{3}$$

with α denoting the complex amplitude accounting for both the target as well as the channel propagation effects and p is the target space-time steering vector. More precisely, the vector p of the returns from a target with normalized Doppler frequency ω_{tg} and spatial frequency θ_{tg} can be written as

$$\boldsymbol{p} = \boldsymbol{a}(\omega_{tg}) \otimes \boldsymbol{b}(\theta_{tg}) \qquad (NL - \text{dimensional}) \tag{4}$$

where $\otimes$ denotes the Kronecker product,

$$\boldsymbol{b}(\theta_{tg}) = [1, \exp(j2\pi\theta_{tg}), \ldots, \exp(j2\pi\theta_{tg}(L-1))]^T \qquad (L - \text{dimensional}), \tag{5}$$

and

$$\boldsymbol{a}(\omega_{tg}) = [1, \exp(j2\pi\omega_{tg}), \ldots, \exp(j2\pi\omega_{tg}(N-1))]^T \qquad (N-\text{dimensional}), \tag{6}$$

with $j = \sqrt{-1}$.

In the sequel we assume that the disturbance vector *d* is a zero-mean complex circular Gaussian vector with covariance matrix

$$\boldsymbol{R} = \boldsymbol{R}_c + \sigma^2 \boldsymbol{I} \tag{7}$$

where R_c denotes the clutter covariance matrix, σ^2 the thermal noise level, and *I* is the identity matrix.

3 On the Structure of the Clutter Covariance Matrix

According to [1], neglecting the ICM, the clutter covariance matrix can be written as

$$\boldsymbol{R}_c = \sum_{i=1}^{N_c} \psi_i \; \boldsymbol{u}(\theta_i, \omega_i) \; \boldsymbol{u}^\dagger(\theta_i, \omega_i) \tag{8}$$

where $(.)^\dagger$ is the conjugate transpose operator, N_c denotes the number of clutter patches involved in the scattering mechanism, φ_i is a positive real number accounting for the power reflected by the *i*-th clutter patch, θ_i and ω_i denote respectively the spatial frequency and the normalized Doppler frequency of the *i*-th clutter patch, and

$$\boldsymbol{u}(\theta_i, \omega_i) = \boldsymbol{a}(\omega_i) \otimes \boldsymbol{b}(\theta_i) . \tag{9}$$

Moreover, since in the absence of velocity misalignment

$$\omega_i = \frac{2v_a T_p}{d}\theta_i = \beta\theta_i \qquad \forall i = 1, \ldots, N_c, \tag{10}$$

with υ_a the platform velocity, we can rewrite R_c as

$$\boldsymbol{R}_c = \sum_{i=1}^{N_c} \psi_i \; \boldsymbol{v}(\theta_i) \; \boldsymbol{v}^\dagger(\theta_i) \tag{11}$$

where $\upsilon(\theta_i) = \upsilon(\theta_i, \beta\theta_i)$. In the sequel we suppose that β is an integer, so that the Brennan rule holds with equality [1], i.e.

$$\text{rank}(\boldsymbol{R}_c) = \min(r, N_c) , \tag{12}$$

with $r = (N-1)\beta + L$, and develop an alternative expression for R_c. To this end we first prove the following proposition.

Proposition 1: Denote by $\overline{\theta}_1, \ldots, \overline{\theta}_r$, r-distinct real values complying with the condition $\overline{\theta}_l \in [0, 1[$, $l = 1, \ldots, r$. Then $\forall i \in \{1, \ldots, N_c\}$

$$\boldsymbol{v}(\theta_i) = \sum_{l=1}^{r} a_{i,l}\, \boldsymbol{v}(\overline{\theta}_l) \qquad a_{i,l} \in \mathcal{C}\,. \tag{13}$$

Proof: Consider the $NL \times (r+1)$ matrix

$$\boldsymbol{V} = \left[\, \boldsymbol{v}(\overline{\theta}_1), \ldots,\ \boldsymbol{v}(\overline{\theta}_r),\ \boldsymbol{v}(\theta_i) \right] \tag{14}$$

with $i \in \{1,\ldots,N_c\}$ and recast it as

$$\boldsymbol{V} = \begin{bmatrix} 1 & \cdots & 1 & 1 \\ \overline{z}_1 & \cdots & \overline{z}_r & z_i \\ \vdots & \vdots & \vdots & \vdots \\ \overline{z}_1^{L-1} & \cdots & \overline{z}_r^{L-1} & z_i^{L-1} \\ \overline{z}_1^{\beta} & \cdots & \overline{z}_r^{\beta} & z_i^{\beta} \\ \vdots & \vdots & \vdots & \vdots \\ \overline{z}_1^{\beta+L-1} & \cdots & \overline{z}_r^{\beta+L-1} & z_i^{\beta+L-1} \\ \overline{z}_1^{(N-1)\beta} & \cdots & \overline{z}_r^{(N-1)\beta} & z_i^{(N-1)\beta} \\ \vdots & \vdots & \vdots & \vdots \\ \overline{z}_1^{(N-1)\beta+L-1} & \cdots & \overline{z}_r^{(N-1)\beta+L-1} & z_i^{(N-1)\beta+L-1} \end{bmatrix} \tag{15}$$

where $\overline{z}_l = \exp(j2\pi\ \overline{\theta}_l)$, l=1, … ,r, and $z_i = \exp(j2\pi\theta_i)$.

Moreover, since the rank of **V** is invariant under permutation of its rows, rearrange them so that the distinct elements of each column occupy the top rows and the replicated entries occupy the bottom rows. This procedure leads to the new matrix whose rank, due to the replicated rows, is at most $r = (N-1)\beta + L$.

$$\boldsymbol{V}_1 = \begin{bmatrix} 1 & \cdots & 1 & 1 \\ \overline{z}_1 & \cdots & \overline{z}_r & z_i \\ \overline{z}_1^2 & \cdots & \overline{z}_r^2 & z_i^2 \\ \vdots & \vdots & \vdots & \vdots \\ \overline{z}_1^{(N-1)\beta+L-1} & \cdots & \overline{z}_r^{(N-1)\beta+L-1} & z_i^{(N-1)\beta+L-1} \\ \vdots & \vdots & \vdots & \vdots \\ & \text{repeated} & \text{rows} & \end{bmatrix} \tag{16}$$

Finally, since its first r columns are linearly independent, due to the assumption on the $\overline{\theta}_i$'s, then rank(**V**) = rank($\mathbf{V}_1$) = r and the (r + 1)-th column can be expressed as a linear combination of the first r.

Exploiting the result of the above proposition we can rewrite $\boldsymbol{R}_c$ as follows

$$\boldsymbol{R}_c = \sum_{i=1}^{N_c} \psi_i \left(\sum_{l=1}^{r} a_{i,l}\ \boldsymbol{v}(\bar{\theta}_l) \right) \left(\sum_{t=1}^{r} a_{i,t}\ \boldsymbol{v}(\bar{\theta}_t) \right)^{\dagger} \tag{17}$$

or equivalently as,

$$\boldsymbol{R}_c = \sum_{l=1}^{r} \sum_{t=1}^{r} \sum_{i=1}^{N_c} \psi_i a_{i,l} a_{i,t}^{*}\ \boldsymbol{v}(\bar{\theta}_l)\ \boldsymbol{v}^{\dagger}(\bar{\theta}_t) = \sum_{l=1}^{r} \sum_{t=1}^{r} s_{l,t}\ \boldsymbol{v}(\bar{\theta}_l)\ \boldsymbol{v}^{\dagger}(\bar{\theta}_t) \tag{18}$$

where $(\cdot)^*$ is the complex conjugate operator and

$$s_{l,t} = \sum_{i=1}^{N_c} \psi_i a_{i,l} a_{i,t}^{*}\,. \tag{19}$$

Finally, in matrix form, $\boldsymbol{R}_c$ can be compactly written as

$$R_c = ASA^{\dagger} \tag{20}$$

Where $A = \left[v(\bar{\theta})_1, ..., v(\bar{\theta})_r \right]$, and $\boldsymbol{S}$ is a non-negative definite $r \times r$ hermitian matrix. We stress the importance and originality of the representation (20) which, due to the result of Proposition 1, is independent of the particular set $(\bar{\theta})_1, \ldots, (\bar{\theta})_r$ chosen. In other words, when β is an integer, the clutter subspace is known and coincides with the subspace spanned by the columns of $\boldsymbol{A}$ irrespective of the particular choice concerning $(\bar{\theta})_1, \ldots, (\bar{\theta})_r$.

4. Derivation of the KB-GLRT

This section is devoted to the derivation of the GLRT for the hypothesis testing problem of Section 2. To this end, as in [20], we assume that a set of secondary data $\boldsymbol{r_1}, \ldots, \boldsymbol{r_K}$ $(K \geq NL)$, namely vectors from range cells surrounding the one being tested, free of useful signal components, and with the same spectral properties of $\boldsymbol{r}$, is available.[2] Then we consider the following GLRT decision rule

$$\frac{\max_{\Theta_1} f(\boldsymbol{r},\ \boldsymbol{r}_1, \ldots,\ \boldsymbol{r}_K | \Theta_1, H_1)}{\max_{\Theta_0} f(\boldsymbol{r},\ \boldsymbol{r}_1, \ldots,\ \boldsymbol{r}_K | \Theta_0, H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} G \tag{21}$$

where Θ_0 is the set of unknown parameters under the hypothesis H_0, $\Theta_1 = \Theta_0 \cup \{\alpha\}$ is the set of unknown parameters under the hypothesis H_1, $f(\boldsymbol{r}, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_K | \Theta_1, H_1)$ and $f(\boldsymbol{r}, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_K | \Theta_0, H_0)$ denote the probability density functions (pdf's) of the data under the H_1 and H_0 hypotheses

$$\begin{cases} \Theta_0 = \{\boldsymbol{S}, \sigma^2\}, & \boldsymbol{S} \geq 0 \\ \Theta_1 = \{\alpha, \boldsymbol{S}, \sigma^2\}, & \boldsymbol{S} \geq 0 \end{cases} \tag{22}$$

the GLRT can be rewritten as

$$\frac{\max_{\alpha, \boldsymbol{S}, \sigma^2} f(\boldsymbol{r},\ \boldsymbol{r}_1, \ldots,\ \boldsymbol{r}_K | \alpha, \boldsymbol{S}, \sigma^2, H_1)}{\max_{\boldsymbol{S}, \sigma^2} f(\boldsymbol{r},\ \boldsymbol{r}_1, \ldots,\ \boldsymbol{r}_K | \boldsymbol{S}, \sigma^2, H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} G\,. \tag{23}$$

[2]The secondary data are modeled as independent and identically distributed zero-mean complex Gaussian random vectors.

respectively. Since

Further developments require specifying the data pd's under both hypotheses. To this end we

$$f(\boldsymbol{r},\ \boldsymbol{r}_1,\ldots,\ \boldsymbol{r}_K|\alpha,\boldsymbol{S},\sigma^2,H_1)=$$

$$\left\{\frac{1}{\pi^N\det(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^\dagger+\sigma^2\boldsymbol{I})}e^{-\mathrm{tr}\left[(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^\dagger+\sigma^2\boldsymbol{I})^{-1}\boldsymbol{T}_1(\alpha)\right]}\right\}^{K+1} \tag{24}$$

and

$$f(\boldsymbol{r},\ \boldsymbol{r}_1,\ldots,\ \boldsymbol{r}_K|\boldsymbol{S},\sigma^2,H_0)=$$

$$\left\{\frac{1}{\pi^N\det(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^\dagger+\sigma^2\boldsymbol{I})}e^{-\mathrm{tr}\left[(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^\dagger+\sigma^2\boldsymbol{I})^{-1}\boldsymbol{T}_0\right]}\right\}^{K+1} \tag{25}$$

highlight that

where det(·) and tr(·) denote respectively the determinant and the trace of a square matrix, while

$$\begin{cases}\boldsymbol{T}_1(\alpha)=\dfrac{1}{K+1}\left[(\boldsymbol{r}-\alpha\boldsymbol{p})(\boldsymbol{r}-\alpha\boldsymbol{p})^\dagger+\displaystyle\sum_{k=1}^{K}\boldsymbol{r}_k\,\boldsymbol{r}_k^\dagger\right]\\ \boldsymbol{T}_0=\dfrac{1}{K+1}\left(\boldsymbol{r}\,\boldsymbol{r}^\dagger+\displaystyle\sum_{k=1}^{K}\boldsymbol{r}_k\,\boldsymbol{r}_k^\dagger\right)\end{cases}$$

$\boldsymbol{T}_1$ and $\boldsymbol{T}_0$ are defined as

The maximization of the denominator of (23) can be performed exploiting the result of [21].

$$\widehat{\boldsymbol{R}}_{0,ML}=\ \mathbf{Proc}(\boldsymbol{T}_0) \tag{26}$$

Precisely, the Maximum Likelihood (ML) estimator of $\boldsymbol{R}$ under the H_0 hypothesis is given by where the procedure **Proc** (·) is defined in Appendix A.

It is now necessary to perform the maximization of the numerator. To this end we observe that maximizing (24) over α, for known $\boldsymbol{R}$, gives the following equation

$$\alpha=\frac{\boldsymbol{p}^\dagger\boldsymbol{R}^{-1}\,\boldsymbol{r}}{\boldsymbol{p}^\dagger\boldsymbol{R}^{-1}\boldsymbol{p}} \tag{27}$$

Moreover, maximizing (24) over $\boldsymbol{R}$, for known α we get

$$\boldsymbol{R}=\ \mathbf{Proc}(\boldsymbol{T}_1(\alpha))\,. \tag{28}$$

It follows that (27) and (28) are two sets of conditions which implicitly define the joint ML estimates of α and $\boldsymbol{R}$. Unfortunately, it seems that the values of α and $\boldsymbol{R}$ capable of jointly satisfying the two conditions cannot be expressed in an explicit form.

However, it is possible to define a recursive procedure, based on cyclic maximizations of (24), capable of increasing an initial value of the likelihood function. Precisely let $\hat{\boldsymbol{R}}^{(t)}$ and $\hat{\alpha}^{(t)}$ be the estimates of $\boldsymbol{R}$ and α at the *t*-th step of the recursion and evaluate.

$$\widehat{\alpha}^{(t+1)} = \text{argmax}_{\alpha} f(\ \boldsymbol{r},\ \boldsymbol{r}_1, \ldots,\ \boldsymbol{r}_K | \alpha, \widehat{\boldsymbol{R}}^{(t)}, H_1) = \frac{\boldsymbol{p}^{\dagger}(\widehat{\boldsymbol{R}}^{(t)})^{-1}\ \boldsymbol{r}}{\boldsymbol{p}^{\dagger}(\widehat{\boldsymbol{R}}^{(t)})^{-1}\boldsymbol{p}} \tag{29}$$

and

$$\widehat{\boldsymbol{R}}^{(t+1)} = \text{argmax}_{\boldsymbol{R}} f(\ \boldsymbol{r},\ \boldsymbol{r}_1, \ldots,\ \boldsymbol{r}_K | \widehat{\alpha}^{(t+1)}, \boldsymbol{R}, H_1) = \ \mathbf{Proc}\left(\boldsymbol{T}_1\left(\widehat{\alpha}^{(t+1)}\right)\right) \tag{30}$$

where $\text{argmax}_{\chi}(\cdot)$ denotes the value of χ which maximizes the argument.

The recursion is such that the sequence of the values of the likelihood function, namely the sequence

$$f\left(\ \boldsymbol{r},\ \boldsymbol{r}_1, \ldots,\ \boldsymbol{r}_K | \widehat{\alpha}^{(t)}, \widehat{\boldsymbol{R}}^{(t)}, H_1\right) \quad ,$$

increases **as *t* increases.** Precisely, the following statement holds true.

Proposition 2: The sequence of the values of the likelihood function is increasing, namely: Not only the likelihood sequence is increasing, but it is also convergent, namely

$$f\left(\ \boldsymbol{r},\ \boldsymbol{r}_1, \ldots,\ \boldsymbol{r}_K | \widehat{\alpha}^{(t)}, \widehat{\boldsymbol{R}}^{(t)}, H_1\right) \leq f\left(\ \boldsymbol{r},\ \boldsymbol{r}_1, \ldots,\ \boldsymbol{r}_K | \widehat{\alpha}^{(t+1)}, \widehat{\boldsymbol{R}}^{(t+1)}, H_1\right) \tag{31}$$

Proof: In order to prove equation (31) it is sufficient to observe that

$$\begin{aligned} f\left(\ \boldsymbol{r},\ \boldsymbol{r}_1, \ldots,\ \boldsymbol{r}_K | \widehat{\alpha}^{(t)}, \widehat{\boldsymbol{R}}^{(t)}, H_1\right) &\leq \max_{\alpha} f\left(\ \boldsymbol{r},\ \boldsymbol{r}_1, \ldots,\ \boldsymbol{r}_K | \alpha, \widehat{\boldsymbol{R}}^{(t)}, H_1\right) \\ &= f\left(\ \boldsymbol{r},\ \boldsymbol{r}_1, \ldots,\ \boldsymbol{r}_K | \widehat{\alpha}^{(t+1)}, \widehat{\boldsymbol{R}}^{(t)}, H_1\right) \\ &\leq \max_{\boldsymbol{R}} f\left(\ \boldsymbol{r},\ \boldsymbol{r}_1, \ldots,\ \boldsymbol{r}_K | \widehat{\alpha}^{(t+1)}, \boldsymbol{R}, H_1\right) \\ &= f\left(\ \boldsymbol{r},\ \boldsymbol{r}_1, \ldots,\ \boldsymbol{r}_K | \widehat{\alpha}^{(t+1)}, \widehat{\boldsymbol{R}}^{(t+1)}, H_1\right) \end{aligned}$$

■

Not only is the likelihood sequence increasing, but it is also convergent, namely

$$\lim_{t} f\left(\ \boldsymbol{r},\ \boldsymbol{r}_1, \ldots,\ \boldsymbol{r}_K | \widehat{\alpha}^{(t)}, \widehat{\boldsymbol{R}}^{(t)}, H_1\right) < +\infty. \tag{32}$$

In fact, given ***r,*** $\boldsymbol{r}_1$, …, $\boldsymbol{r}_K$, the likelihood function is upper-bounded.

Starting from the above considerations we obtain the following recursive expression of the KB-GLRT

$$\frac{f(\ \boldsymbol{r},\ \boldsymbol{r}_1, \ldots,\ \boldsymbol{r}_K | \widehat{\alpha}^{(t)}, \widehat{\boldsymbol{R}}^{(t)}, H_1)}{f(\ \boldsymbol{r},\ \boldsymbol{r}_1, \ldots,\ \boldsymbol{r}_K | \widehat{\boldsymbol{R}}_{0,ML}, H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} G, \tag{33}$$

which taking the logarithm can be recast as

$$\log\left[\frac{\det\left(\widehat{\boldsymbol{R}}_{0,ML}\right)}{\det\left(\widehat{\boldsymbol{R}}^{(t)}\right)}\right]-\mathrm{tr}\left[\left(\widehat{\boldsymbol{R}}^{(t)}\right)^{-1}\boldsymbol{T}_1\left(\widehat{\alpha}^{(t)}\right)-\left(\widehat{\boldsymbol{R}}_{0,ML}\right)^{-1}\boldsymbol{T}_0\right]\underset{H_0}{\overset{H_1}{\gtrless}}G, \qquad (34)$$

where G denotes a suitable modification of the original threshold.

For its implementation only an initial estimate $\boldsymbol{R}$ is required. To this end it is convenient to exploit the ML estimate of $\boldsymbol{R}$ evaluated from the secondary data, i.e.

$$\widehat{\boldsymbol{R}}^{(0)}=\widehat{\boldsymbol{R}}_s=\mathbf{Proc}\left(\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{r}_k\,\boldsymbol{r}_k^{\dagger}\right). \qquad (35)$$

A summary of the complete algorithm for the evaluation of the KB-GLRT (34) is given below.

1. Evaluate (35).
2. For t=1, …, N_{it} compute $\hat{\alpha}^{(t)}$ $\hat{\boldsymbol{R}}^{(t)}$.
3. Perform the test (34).

Finally, considerations concerning the mathematical tractability of the problem for non-integer values of β are given in Section 7.

5. Derivation of the KB-2S-GLRT

In this section we devise the 2S-GLRT detector [22] exploiting the *a-priori* information about the coupling between the normalized Doppler and the spatial frequencies. Otherwise stated, we first derive the GLRT based upon the primary data vector $\boldsymbol{r}$ assuming that the covariance matrix $\boldsymbol{R}$ is known. Then a fully adaptive detector is obtained substituting the unknown matrix by its ML estimate based upon secondary data only.

Step 1. For known $\boldsymbol{R}$ the GLRT based upon the primary data is the following decision rule

$$\frac{|\boldsymbol{p}^{\dagger}\boldsymbol{R}^{-1}\,\boldsymbol{r}|^2}{\boldsymbol{p}^{\dagger}\boldsymbol{R}^{-1}\boldsymbol{p}}\underset{H_0}{\overset{H_1}{\gtrless}}G. \qquad (36)$$

Step 2. We can make detector (36) fully adaptive by plugging the ML estimate of $\boldsymbol{R}$, based upon the secondary data, in place of its exact value. By doing so we come up with the following adaptive decision rule (KB-2S-GLRT)

$$\frac{|\boldsymbol{p}^{\dagger}\widehat{\boldsymbol{R}}_s^{-1}\,\boldsymbol{r}|^2}{\boldsymbol{p}^{\dagger}\widehat{\boldsymbol{R}}_s^{-1}\boldsymbol{p}}\underset{H_0}{\overset{H_1}{\gtrless}}G. \qquad (37)$$

In Section 7 some considerations for an approximate implementation of the receiver for the case of a non-integer clutter ridge slope are provided.

6. Performance Assessment

This section is devoted to the performance assessment of the KB-GLRT and of the KB-2S-GLRT also in comparison with

- the optimum receiver (for targets with uniformly distributed phase) which assumes perfect knowledge of the clutter covariance matrix, i.e.,

$$|\boldsymbol{p}^{\dagger}\boldsymbol{R}^{-1}\,\boldsymbol{r}|^2 \underset{H_0}{\overset{H_1}{\gtrless}} G \qquad \text{Optimum-RX}. \tag{38}$$

- classical Kelly's receiver and the Adaptive Matched Filter (AMF) which, in spite of the new detectors, do not exploit the special structure of the clutter PSD (the coupling between Doppler and spatial frequency), i.e.

$$\frac{|\boldsymbol{p}^{\dagger}\boldsymbol{S}_s^{-1}\,\boldsymbol{r}|^2}{\left(1+\frac{1}{K}\,\boldsymbol{r}^{\dagger}\boldsymbol{S}_s^{-1}\,\boldsymbol{r}\right)\boldsymbol{p}^{\dagger}\boldsymbol{S}_s^{-1}\boldsymbol{p}} \underset{H_0}{\overset{H_1}{\gtrless}} G \qquad \text{Kelly-RX}, \tag{39}$$

and

$$\frac{|\boldsymbol{p}^{\dagger}\boldsymbol{S}_s^{-1}\,\boldsymbol{r}|^2}{\boldsymbol{p}^{\dagger}\boldsymbol{S}_s^{-1}\boldsymbol{p}} \underset{H_0}{\overset{H_1}{\gtrless}} G \qquad \text{AMF-RX}, \tag{40}$$

where

$$\boldsymbol{S}_s = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{r}_k\,\boldsymbol{r}_k^{\dagger},$$

is the secondary data sample covariance matrix.

To this end we resort to Monte Carlo simulation based on $\frac{100}{P_{fa}}$ independent trials. Moreover, in order to limit the computational burden we set $P_{fa} = 10^{-2}$. We also assume $L = 4$, $N = 6$, $\theta_{tg} = 0$, $\omega_{tg} = 0.01$, and $N_c = 1000$ independent and identically distributed clutter patches. The Clutter-to-Noise Ratio (CNR) is defined as

$$CNR = \frac{\mathrm{tr}(\boldsymbol{R}_c)}{\sigma^2}, \tag{41}$$

and is chosen equal to 30 dB. Finally, simulation results have shown that $N_{it} = 1$ is sufficient in order for the KB-GLRT to achieve convergence.

In Figures 1, 2, and 3 the probability of detection (P_d) is plotted versus the Signal to Interference plus Noise Power Ratio (SINR) for $\beta = 1$ and several values of K. SINR is given by:

$$SINR = |\alpha|^2 \boldsymbol{p}^{\dagger}\boldsymbol{R}^{-1}\boldsymbol{p}, \tag{42}$$

where $|\cdot|$ denotes the absolute value of a complex number. The results show that the KB-GLRT outperforms the KB-2S-GLRT. Both new detectors exhibit a P_d level higher than Kelly's receiver and the AMF. Moreover, the performance gain depends on the size of the available sample support. The smaller K is the higher the gain. Finally, for the parameter values chosen, the loss of the new receivers with respect to the optimum processor is acceptable even for small sample supports, and the loss decreases as K increases.

In Figure 4 the effect of Doppler ambiguous clutter is analyzed. We assume $\beta = 2$, $K = 3$-, and the same target parameters of the previous three figures. The curves show that even in the presence of an aliased clutter ridge the novel receivers outperform Kelly's detector and the AMF, especially for values of P_d in the low-medium range.

7. Analytical Problems Connected with the Case of a Non-Integer β and a Possible Approximate Solution

In this section we highlight the analytical complications which were raised in the derivation of the GLRT with reference to the case of a non-integer value of the parameter β, and we propose two approximate decision strategies.

To this end we observe that according to the Brennan rule the rank of the clutter covariance matrix can be approximated as [1]

$$\text{rank}(\boldsymbol{R}_c) = \min(r, N_c)\ , \tag{43}$$

Where r = round $((N-1)\,\beta + L)$, and round $(\cdot)$ indicates rounding to the nearest integer. As a consequence the clutter covariance matrix still admits the representation in (20), but Proposition 1 does not hold true. Otherwise stated, the values $\overline{\theta}_1, \ldots, \overline{\theta}_r$ cannot be arbitrarily chosen and the matrix $\boldsymbol{A}$ must be considered a function of the vector $\theta = [(\overline{\theta})_1, \ldots, (\overline{\theta})_r]$, i.e. $\boldsymbol{A} = \boldsymbol{A}(\boldsymbol{\theta})$. It follows that the parameter sets under the H_0 and H_1 hypothesis are

$$\begin{cases} \Theta_0 = \{\boldsymbol{\theta}, \boldsymbol{S}, \sigma^2\}\,, & \boldsymbol{S} \geq 0 \quad \boldsymbol{\theta} \in [0,1[^r \\ \Theta_1 = \{\alpha, \boldsymbol{\theta}, \boldsymbol{S}, \sigma^2\}\,, & \boldsymbol{S} \geq 0 \quad \boldsymbol{\theta} \in [0,1[^r \end{cases} \tag{44}$$

In order to evaluate the GLRT it is necessary to calculate the ML estimates of the parameters under both hypotheses. Thus we first focus on the H_1 hypothesis and consider the problem

$$\max_{\alpha, \boldsymbol{\theta}, \boldsymbol{S}, \sigma^2} f(\boldsymbol{r},\ \boldsymbol{r}_1, \ldots,\ \boldsymbol{r}_K | \alpha, \boldsymbol{\theta}, \boldsymbol{S}, \sigma^2, H_1)\,. \tag{45}$$

Maximizing over α, for fixed $\boldsymbol{R}$, yields

$$\alpha(\boldsymbol{\theta}, \boldsymbol{R}) = \frac{\boldsymbol{p}^\dagger \boldsymbol{R}^{-1}\, \boldsymbol{r}}{\boldsymbol{p}^\dagger \boldsymbol{R}^{-1} \boldsymbol{p}}\,. \tag{46}$$

Moreover, maximizing over $\boldsymbol{S} \geq 0$ and σ^2 for fixed α and $\boldsymbol{\theta}$ we get

$$\boldsymbol{R}(\alpha, \boldsymbol{\theta}) = \mathbf{Proc}(\boldsymbol{T}_1(\alpha)) . \tag{47}$$

Finally, differentiating $f(\boldsymbol{r}, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_K | \alpha, \boldsymbol{\theta}, \boldsymbol{S}, \sigma^2, H_1)$ with respect to the components of $\boldsymbol{\theta}$, for fixed α, $\boldsymbol{S}$, and σ^2, we come up with the following set of non-linear equations which implicitly define a stationary point:

$$\begin{aligned} &\mathrm{tr}\left\{\left[\left(\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{S}\boldsymbol{A}^{\dagger}(\boldsymbol{\theta}) + \sigma^2\boldsymbol{I}\right)^{-1}\left(\frac{\partial \boldsymbol{A}(\boldsymbol{\theta})}{\partial \bar{\theta}_i}\boldsymbol{S}\boldsymbol{A}^{\dagger}(\boldsymbol{\theta}) + \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{S}\frac{\partial \boldsymbol{A}^{\dagger}(\boldsymbol{\theta})}{\partial \bar{\theta}_i}\right)\right] \times \right. \\ &\left.\left[\left(\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{S}\boldsymbol{A}^{\dagger}(\boldsymbol{\theta}) + \sigma^2\boldsymbol{I}\right)^{-1}\boldsymbol{T}_1(\alpha) - \boldsymbol{I}\right]\right\} = 0 \qquad i = 1, \ldots, r , \end{aligned} \tag{48}$$

where

$$\frac{\partial \boldsymbol{A}(\boldsymbol{\theta})}{\partial \bar{\theta}_i} = \frac{\partial\, \boldsymbol{v}(\bar{\theta}_i)}{\partial \bar{\theta}_i}\boldsymbol{e}_i^T , \tag{49}$$

and e_i is the NL column vector with all zero components but for the i-th which is equal to 1.

The joint ML estimates of the parameters must satisfy jointly (46), (47), and (48). This is an apparently mathematical intractable problem. Similar considerations can also be done with reference to the maximization problem under the H_0 hypothesis as well as to the derivation of the KB-2S-GLRT for the case of a non-integer beta.

In order to circumvent this drawback we notice that the GLRT strategy (and the two-step GLRT technique) can be still applied in an approximate fashion if we assume that an estimate of $\boldsymbol{A}$ is available. To this end a very simple estimate which could be adopted is represented by the dominant eigenvectors of the secondary data sample covariance matrix. We explicitly point out that this solution does not sensibly increase the computational burden connected with the evaluation of the decision rule since the eigenvalue decomposition of the aforementioned matrix must be evaluated as required by (35).

The receivers which exploit the quoted approximation will be referred to, in the sequel, as the Approximate KB-GLRT (AKB-GLRT) and the approximate KB-2S-GLRT (AKB-2S-GLRT). Their detection performance (evaluated by Monte Carlo simulation) is analyzed in Figures 5 and 6 also in comparison with the optimum receiver, the Kelly receiver, and the AMF for $P_{fa} = 10^{-2}$, $L = 4$, $N = 6$, $K = 3$-, $\theta_{tg} = 0$, $\omega_{tg} = 0.01$, and $N_c = 1000$ independent and identically distributed clutter patches. Figure 5 refers to $\beta = 0.9$ whereas Figure 6 assumes $\beta = 1.1$. Finally we assume $N_{it} = 1$, and $CNR = 30$ dB.

The curves show that in both situations the AKB-GLRT and the AKB-2S-GLRT outperform the Kelly receiver and the AMF. In particular, for $P_d = 0.9$, the performance gain is kept within 5 dB for both $\beta = 0.9$ and $\beta = 1.1$. Finally, always for $P_d = 0.9$, the loss of the knowledge-based receivers with respect to the perfect measurement bound is about 4 db (if $\beta = 0.9$) or 5 dB (if $\beta = 1.1$).

Notice that a different technique to cope with the uncertainty about the clutter subspace might be the employment of further *a-priori* information. This is tantamount to assuming the existence of a database containing the structure of the covariance subspace for several different β values. Based on this knowledge the processor first selects the actual environment (and as a consequence the specific clutter subspace) and then evaluates the decision statistics for testing the target presence.

8. Conclusions

In this paper we have proposed and assessed knowledge-based detectors for STAP applications. Exploiting the special structure of the clutter ridge we have derived a relevant property of the clutter space-time covariance matrix. We have shown that for integer values of the slope parameter the clutter covariance subspace is perfectly known. Thus we have devised the KB-GLRT and the KB-2S-GLRT which exploit the aforementioned property. We have also addressed the most general case of clutter ridge with non-integer slope and have shown that the problem is very difficult from an analytical point of view since it involves the solution of multivariate non-linear equations. Approximate GLRT detectors have been thus proposed and hints for their implementation with a supporting environmental database have been provided.

The performance analysis has shown that the new knowledge-based systems achieve a performance level very close to the optimum detector which assumes perfect knowledge of the clutter covariance matrix. Finally, the new receivers can also outperform some previously proposed adaptive detection schemes such as the Kelly receiver and the AMF.

References

[1] J. Ward: "Space-Time Adaptive Processing for Airborne Radar," Tech. Report No. 1015, Lincoln Laboratories, December 13, 1994.

[2] R. Klemm: "Principles of Space-Time Adaptive Processing," IEE Press, 2002.

[3] L. E. Brennan and I. S. Reed: "Theory of Adaptive Radar," IEEE Trans. On Aerospace and Electronic Systems, Vol. 9, No. 1, pp. 237-252, March 1973.

[4] I. S. Reed, J. D. Mallett, and L. E. Brennan: "Rapid Convergence Rate in Adaptive Arrays," IEEE Trans. on Aerospace and Electronic Systems, Vol. 10, No. 4, pp. 853- 863, Nov 1974.

[5] J. S. Goldstein, I. S. Reed, and P. A. Zulch: "Multistage Partially Adaptive STAP CFAR Detection Algorithm," IEEE Trans. on Aerospace and Electronic Systems, Vol. 35, No. 2, pp. 645-661, April 1999.

[6] H. Wang and L. Cai: "On the Adaptive Spatial-Temporal Processing for Airborne Surveillance Radar Systems," IEEE Trans. on Aerospace and Electronic Systems, Vol. 30, No. 3, pp. 660-669, July 1994.

[7] J. R. Guerci and J. S. Bergin: "Principal Components, Covariance Matrix Tapers, and the Subspace Leakage Problem," IEEE Trans. on Aerospace and Electronic Systems, Vol. 38, No. 1, pp. 152-162, January 2002.

[8] R. Nitzberg: "Application of Maximum Likelihood Estimation of Persymmetric Covariance Matrices to Adaptive Processing," IEEE Trans. on Aerospace and Electronic Systems, Vol. 16, No. 1, pp. 124-127, January 1980.

[9] T. A. Barton and S. T. Smith: "Structured Covariance Estimation for Space-Time Adaptive Processing," Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 3493-3496, April 1997.

[10] M. Steiner and K. Gerlach: "Fast Converging Adaptive Processors of a Structured Covariance Matrix," IEEE Trans. on Aerospace and Electronic Systems, Vol. 36, No. 4, pp. 1115-1126, October 2000.

[11] J. R. Roman, M. Rangaswamy, D. W. Davis, Q. Zhang, B. Himed, and J. H. Michels: "Parametric Adaptive Matched Filter for Airborne Radar Applications," IEEE Trans. on Aerospace and Electronic Systems, Vol. 36, No. 2, pp. 677-692, April 2000.

[12] A. De Maio and A. Farina: "A Maximum Entropy Framework for Space-Time Adaptive Processing," Signal Processing, pp. 1637-1652, No. 9, Vol. 84, September 2004.

[13] D. J. Rabideau and A. O. Steinhardt: "Improved Adaptive Clutter Cancellation through Data-Adaptive Training," IEEE Trans. on Aerospace and Electronic Systems, Vol. 35, No. 3, pp. 879-891, July 1999.

[14] W. L. Melvin, M. Wicks, P. Antonik, Y. Salama, P. Li, and H. Schuman: "Knowledge-Based Space-Time Adaptive Processing for Airborne Early Warning Radar," IEEE AES Systems Magazine, Vol. , No. , pp. 37-42, April 1998.

[15] R. Adve, P. Antonik, W. Baldygo, C. Capraro, G. Capraro, T. Hale, R. Schneible, and M. Wicks: "Knowledge-base Application to Ground Moving Target Detection," AFRL-SN-TR-2001-185, In-House Technical Report, September 2001.

[16] A. Farina, P. Lombardo, and M. Pirri: "Nonlinear Nonadaptive Space-Time Processing for Airborne Early Warning Radar," IEE Proc. Radar, Sonar Navig., Vol. 145, No. 1, pp. 9-18, February 1998.

[17] A. Farina, P. Lombardo, and M. Pirri: "Nonlinear STAP Processing," ECEJ Special Issue on STAP, Vol. 11, No. 1, pp. 41-48, February 1999.

[18] P. Lombardo and F. Colone: "Non-linear and Adaptive Two-Dimensional FIR Filters for STAP: Theory and Experimental Results," Applications of STAP, edited by R. Klemm, IEE Radar, Sonar and Navigation series 14, pp. 37-72, 2004.

[19] S. Parker and T. K. Sarkar: "A Deterministic Eigenvalue Approach to Space-Time Adaptive Processing," Proc. of IEEE International Symposium on Phased Array Systems and Technology, pp. 372-375, Boston, USA, 1996.

[20] M. Wicks and B. Himed: "Four Problems in Radar," Computational Noncommutative Algebra and Applications, edited by J. Byrnes, NATO Science Series, Vol. 136, pp. 57-73, Kluwer Academic Publishers, 2004.

[21] E. J. Kelly: "An Adaptive Detection Algorithm," IEEE Trans. on Aerospace and Electronic Systems, Vol. 22, No. 1, pp. 115-127, March 1986.

[22] Y. Bresler: "Maximum Likelihood Estimation of a Linearly Structured Covariance with Application to Antenna Array Processing," Proc. of the Fourth Annual ASSP Workshop on Spectral Estimation and Modeling, pp. 172-175, 1988.

[23] F. C. Robey, D. R. Fuhrmann, R. Nitzberg, and E. J. Kelly: "A CFAR Adaptive Matched Filter Detector," IEEE Trans. on Aerospace and Electronic Systems, Vol. 28, No. 1, pp. 208-216, January 1992.

Appendix A: R_{ML} = Proc (T)

9 Appendix A: $\widehat{R}_{ML}$ =Proc(T)

Let

$$L(\boldsymbol{S},\sigma^2)=\left\{\frac{1}{\pi^N\det(\boldsymbol{ASA}^\dagger+\sigma^2\boldsymbol{I})}e^{-\mathrm{tr}\left[(\boldsymbol{ASA}^\dagger+\sigma^2\boldsymbol{I})^{-1}\boldsymbol{T}\right]}\right\}^{K+1}, \tag{50}$$

where $\boldsymbol{T}$ is a positive definite matrix. Then the ML estimator of $\boldsymbol{R}=\boldsymbol{ASA}^\dagger+\sigma^2\boldsymbol{I}$, with the constraint $\boldsymbol{S}\geq 0$, can be evaluated according to the following procedure (**Proc(T)**).

1. Compute the Cholesky factor $\boldsymbol{L}$ of $\boldsymbol{A}^\dagger\boldsymbol{A}$.
2. Form the $r\times r$ matrix $\boldsymbol{X}=\boldsymbol{L}^{-1}\boldsymbol{A}^\dagger\boldsymbol{TAL}^{-\dagger}$ and compute its eigendecomposition $\boldsymbol{X}=\boldsymbol{G\Phi G}^\dagger$, where $\Phi=\mathrm{diag}(\phi_1,\ldots,\phi_r)$, and $\phi_1\geq\ldots\geq\phi_r$.
3. Evaluate $\gamma=\mathrm{tr}(\boldsymbol{P}^\perp\boldsymbol{T})$ where $\boldsymbol{P}^\perp=I-\boldsymbol{A}(\boldsymbol{A}^\dagger\boldsymbol{A})^{-1}\boldsymbol{A}^\dagger$.
4. Set $\widehat{k}=r$.
5. Evaluate $\widehat{\sigma^2}=\frac{1}{NL-\widehat{k}}\left(\gamma+\sum_{i=\widehat{k}+1}^{r}\phi_i\right)$.
6. Verify if[3] $\phi_{\widehat{k}+1}\leq\widehat{\sigma^2}\leq\phi_{\widehat{k}}$. Yes: go to Step 9.
7. $\widehat{k}=\widehat{k}-1$.
8. Go to Step 5.
9. Set $\widehat{\sigma^2}_{ML}=\widehat{\sigma^2}$.
10. Define $\Lambda=\mathrm{diag}(\lambda_1,\ldots,\lambda_r)$ where
$$\lambda_i=\begin{cases}\phi_i-\widehat{\sigma^2} & i=1,\ldots,\widehat{k}\\ 0 & i=\widehat{k}+1,\ldots,r\end{cases}$$
11. $\widehat{\boldsymbol{S}}_{ML}=\boldsymbol{L}^{-\dagger}\boldsymbol{G}\,\Lambda\boldsymbol{G}^\dagger\boldsymbol{L}^{-1}$.
12. $\widehat{\boldsymbol{R}}_{ML}=\boldsymbol{A}\widehat{\boldsymbol{S}}_{ML}\boldsymbol{A}^\dagger+\widehat{\sigma^2}_{ML}\boldsymbol{I}$.

We explicitly point out that the above procedure simplifies if the matrix

$$\boldsymbol{\Lambda}_1=\boldsymbol{\Phi}-\frac{1}{NL-r}\gamma\geq 0. \tag{51}$$

In fact, under this last assumption, $\widehat{\boldsymbol{S}}_{ML}=\boldsymbol{L}^{-\dagger}\boldsymbol{G}\,\boldsymbol{\Lambda}_1\boldsymbol{G}^\dagger\boldsymbol{L}^{-1}$ and $\widehat{\boldsymbol{R}}_{ML}=\boldsymbol{A}\widehat{\boldsymbol{S}}_{ML}\boldsymbol{A}^\dagger+\widehat{\sigma^2}_{ML}\boldsymbol{I}$.

[3] Use the definitions $\phi_0=+\infty$ and $\phi_{r+1}=0$

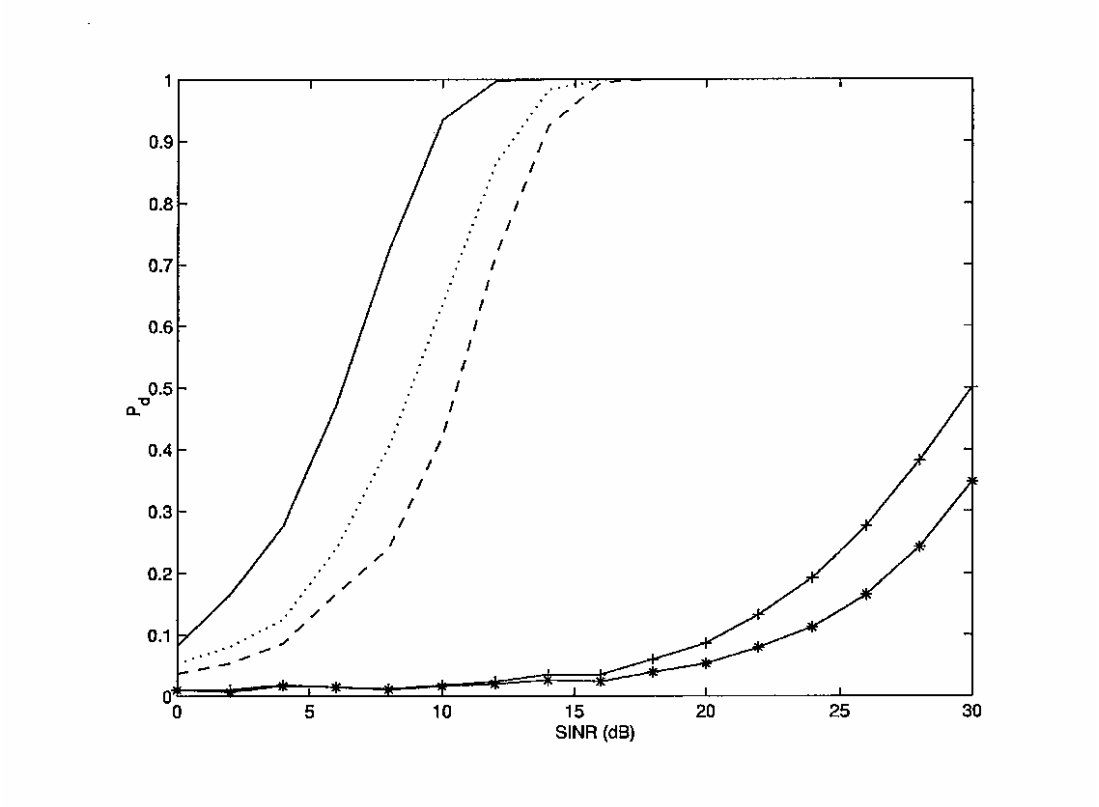

Figure 1: P_d versus *SINR* for $P_{fa} = 10^{-2}$, $L = 4$, $N = 6$, $\theta_{tg} = 0$, $\omega_{tg} = 0.01$, $\beta = 1$, $K = 24$, $N_{it} = 1$, and $CNR = 30$ dB. Optimum receiver (solid curve), KB-GLRT (dotted curve), KB-2S-GLRT (dashed curve), Kelly receiver (plus-marked curve), and AMF (star-marked curve).

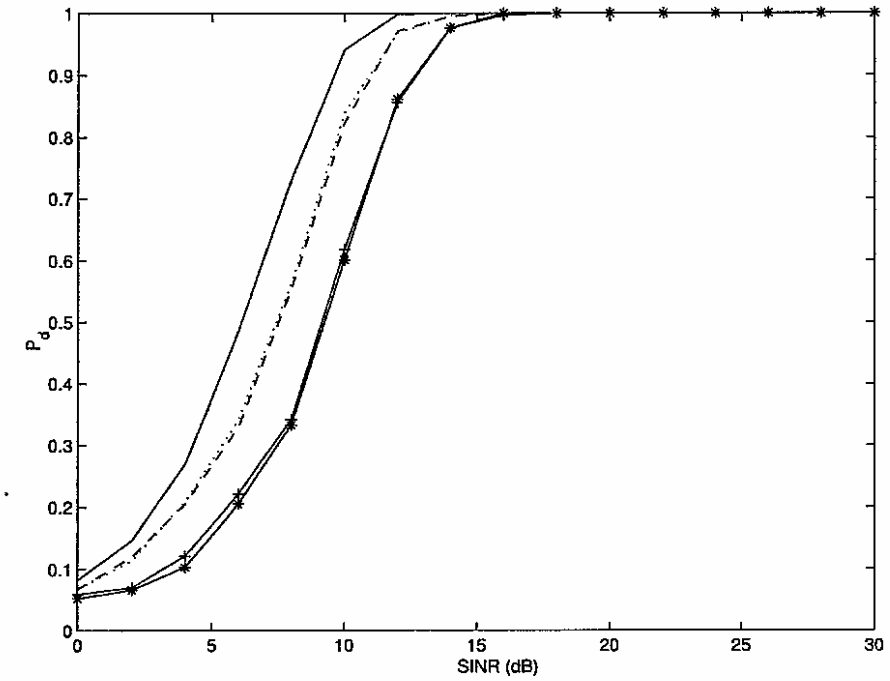

Figure 2: P_d versus *SINR* for $P_{fa} = 10^{-2}$, $L = 4$, $N = 6$, $\theta_{tg} = 0$, $\omega_{tg} = 0.01$, $\beta = 1$, $K = 48$, $N_{it} = 1$, and $CNR = 30$ dB. Optimum receiver (solid curve), KB-GLRT (dotted curve), KB-2S-GLRT (dashed curve), Kelly receiver (plus-marked curve), and AMF (star-marked curve).

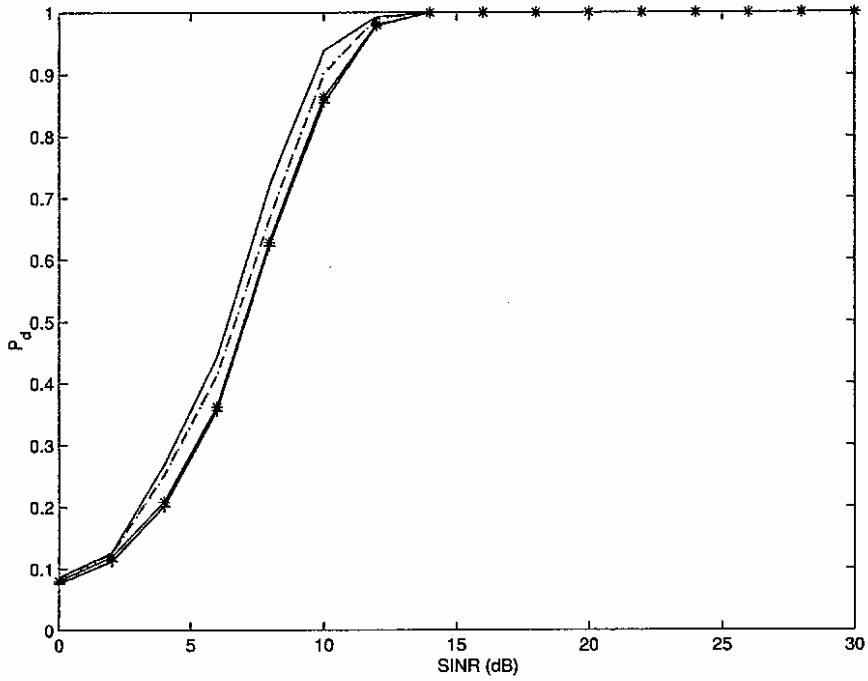

Figure 3: P_d versus *SINR* for $P_{fa} = 10^{-2}$, $L = 4$, $N = 6$, $\theta_{tg} = 0$, $\omega_{tg} = 0.01$, $\beta = 1$, $K = 120$, $N_{it} = 1$, and $CNR = 30$ dB. Optimum receiver (solid curve), KB-GLRT (dotted curve), KB-2S-GLRT (dashed curve), Kelly receiver (plus-marked curve), and AMF (star-marked curve).

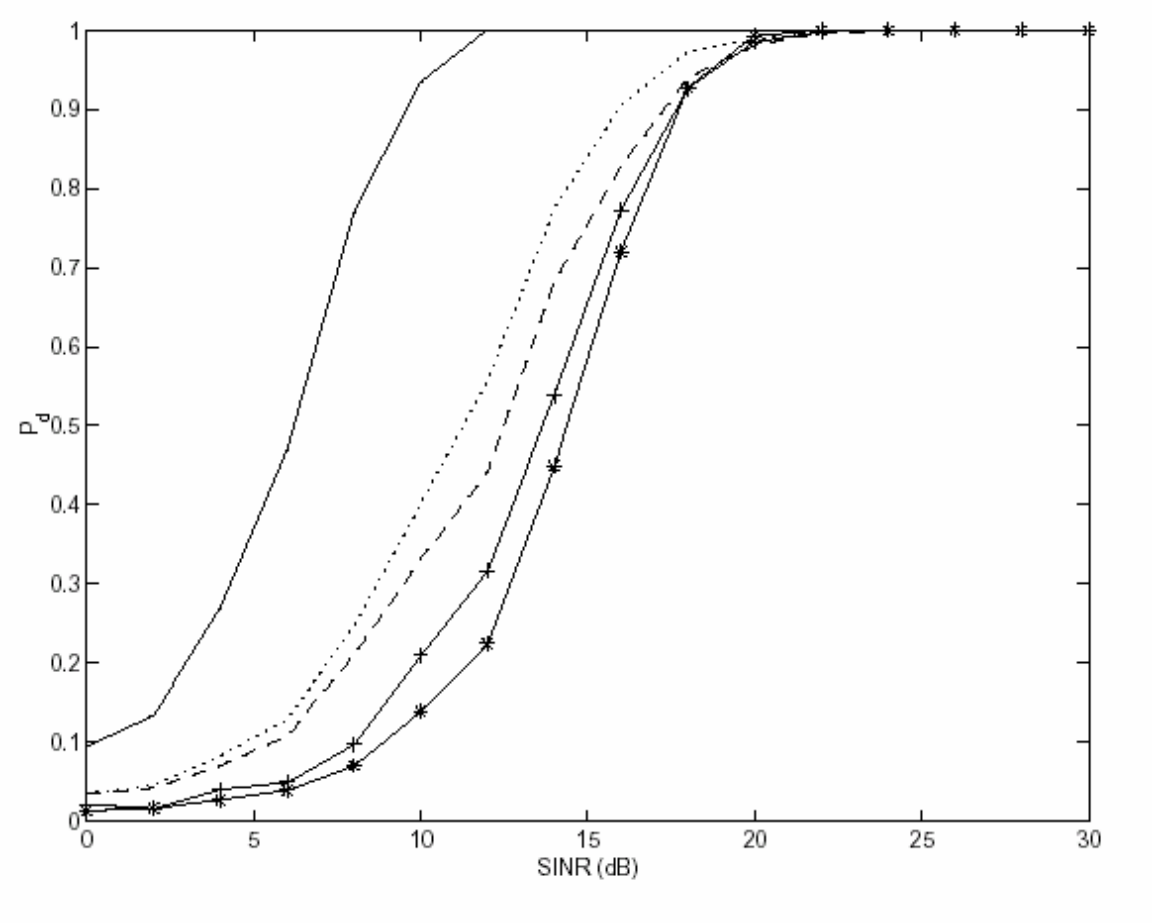

Figure 4: P_d versus *SINR* for $P_{fa} = 10^{-2}$, $L = 4$, $N = 6$, $\theta_{tg} = 0$, $\omega_{tg} = 0.01$, $\beta = 2$, $K = 30$, $N_{it} = 1$, and $CNR = 30$ dB. Optimum receiver (solid curve), KB-GLRT (dotted curve), KB-2S-GLRT (dashed curve), Kelly receiver (plus-marked curve), and AMF (star-marked curve).

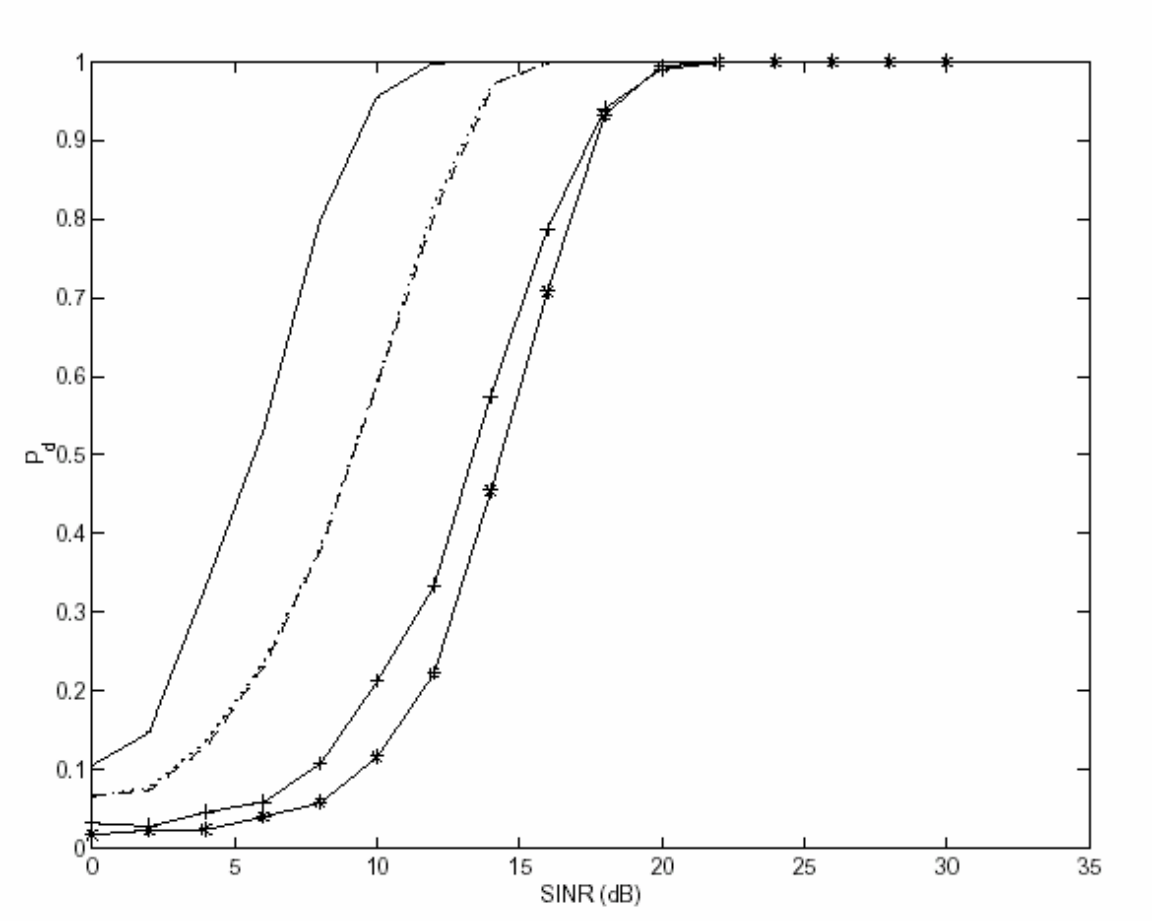

Figure 5: P_d versus $SINR$ for $P_{fa} = 10^{-2}$, $L = 4$, $N = 6$, $\theta_{tg} = 0$, $\omega_{tg} = 0.01$, $\beta = 0.9$, $K = 30$, $N_{it} = 1$, and $CNR = 30$ dB. Optimum receiver (solid curve), AKB-GLRT (dotted curve), AKB-2S-GLRT (dashed curve), Kelly receiver (plus-marked curve), and AMF (star-marked curve).

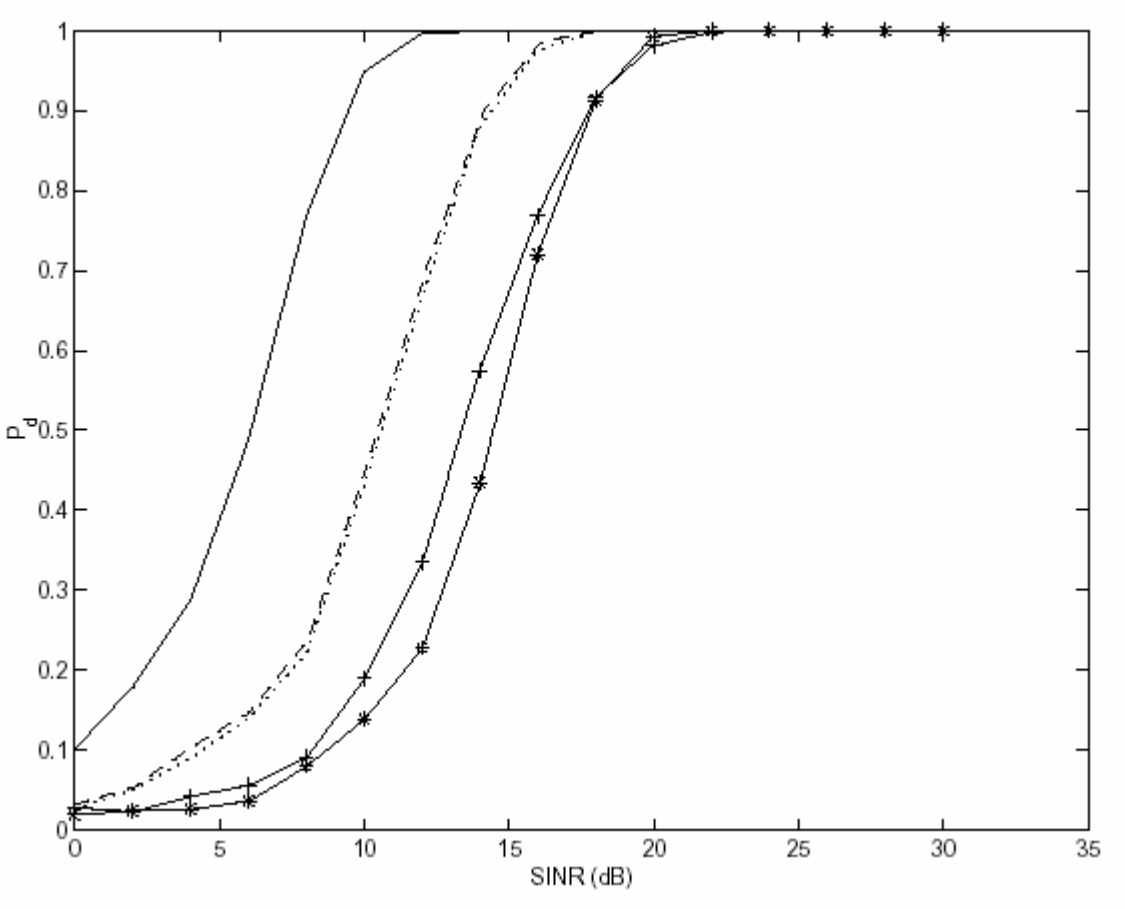

Figure 6: P_d versus *SINR* for $P_{fa} = 10^{-2}$, $L = 4$, $N = 6$, $\theta_{tg} = 0$, $\omega_{tg} = 0.01$, $\beta = 1.1$, $K = 30$, $N_{it} = 1$, and $CNR = 30$ dB. Optimum receiver (solid curve), AKB-GLRT (dotted curve), AKB-2S-GLRT (dashed curve), Kelly receiver (plus-marked curve), and AMF (star-marked curve).